A TOP-LEVEL ONTOLOGY FOR SMART DOCUMENT ACCESS

RONALD MAIER

School of Management, Information Systems
Leopold-Franzens-University of Innsbruck, A-6020 Innsbruck, Austria
E-mail: ronald.maier@uibk.ac.at

JOHANNES SAMETINGER

Inst. f. Wirtschaftsinformatik – Software Engineering,
Johannes Kepler University Linz, A-4040 Linz, Austria
E-mail: johannes.sametinger@jku.at

The central idea of this paper is to apply the concept of a top-level ontology to make access and retrieval of resources, more specifically documents, in knowledge-intensive business processes smarter. Supporting the reuse of meta-data created elsewhere, the sharing of meta-data across applications and users as well as the application of meta-data to answer user enquiries about resources requires a mapping between diverse content-oriented meta-data standards. This can be done with the help of a top-level ontology. In this paper, we will analyze a number of widely used meta-data standards for personal information, communication, resource descriptions and specific standards for text, images, audio and video. These cover a wide range of electronic resources typically required by weakly structured, knowledge-intensive business processes which are then mapped to the presented top-level ontology.

1. Introduction

Businesses and organizations can be characterized by increasingly knowledge-intensive work, activities and business processes [Wolff 2005, 38]. Employees working in knowledge-intensive business processes typically face a multitude of heterogeneous applications with which they access a wealth of electronic resources scattered across a number of systems. These resources are implicitly linked to each other, e.g., by having the same authors, discussing the same topics, using the same formats, referring to the same geographical location or, last but not least, being created, altered or applied in the same processes or projects. Because these links are only implicit, users get no support in retrieving all resources relevant for the tasks they are working on.

Meta-data have been proposed to establish these links and languages have been standardized in order to annotate resources and share meta-data across applications and contexts. Corresponding tools have been developed to support creation, storage and retrieval of meta-data-based descriptions of resources [Maier/Peinl 2005]. However, whereas languages and tools have been standardized, there are a large number of content-oriented meta-data standards for diverse application areas and also annotating resources and thus creating meta-data can become a nuisance to the user.

Section 2 describes the basic terms documents, meta-data and ontology. Section 3 reviews widely used meta-data standards for personal information, communication and resource descriptions. Section 4 shows how they can be mapped to a top-level ontology and discusses how this helps to solve KM challenges. Section 5 briefly compares our approach with related work before Section 6 finally concludes the paper.

2. Document Descriptions

In the following, the three concepts documents, meta-data and ontologies are briefly reviewed from the perspective of ICT environments that provide support for knowledge-intensive work, also called enterprise knowledge infrastructures [Maier et al. 2005].

2.1. *Documents*

Organizations create, store, transfer and use more and more data within their boundaries, but they also exchange data with suppliers, customers and partners. The forms and types of data have been extended largely from structured data found in (relational) data base systems to semi-structured data typically found in file systems, document or content management systems. Compared to structured data, semi-structured data has not been managed equally well in most organizations. A large number of terms have been coined in order to show the various types of semi-structured data, e.g., content, (digital) asset, document, (knowledge, learning or media) object. Documents represent a certain type of container for content, e.g., including formats or meta-data. In the following, we will use the term document as we are interested in the container metaphor. Thus, documents can be exchanged including meta-data so that they can be embedded in a new context at the receiver's end and accessed smartly with the help of the meta-data.

Documents can be searched for with the help of full text search tools included in the operating system or installed on top, e.g., Desktop search engines, search engines for file servers. Meta-data based search, however, is not possible in these cases. When passive documents are uploaded into document management systems (DMS), meta-data search is possible, but in most cases restricted to manually attributed meta-data. Also, when the document is sent to a different environment, it looses all meta-data as these are stored in a DMS data base separated from the documents. This challenge is taken on by the concept of active documents. According to this approach, documents include meta-data and functionality and can be classified into passive, adaptive, active and proactive documents, depending on the amount of functionality that is realized with the help of active components [Schimkat 2003, 54, Carr et al. 2004].

2.2. *Meta-data*

Meta-data is data about data. The term refers to any data used to aid identification, description and location of resources. Examples of meta-data are title, author, date or location used to describe a book or person stored in catalogues, television guides, taxonomies, or tables of contents. Meta-data is an increasingly ubiquitous term that is understood in different ways by different professional communities that design, create, describe, preserve, and use information systems and resources. Until the mid-1990s, meta-data was a term most prevalently used in the domain of geospatial data and data management and systems design and maintenance in general. For these communities, "meta-data" referred to a suite of industry or disciplinary standards as well as additional internal and external documentation and other data necessary for the identification, representation, interoperability, technical management, performance, and use of data contained in an information system [Swetland, Baca 2000]. Many meta-data formats exist, see Section 3, some of these are simple in their descriptions; others are complex and rich.

2.3. *Ontologies*

Knowledge modeling aims at a formal description of (documented) organizational knowledge that can be processed by computers. Ontologies are formal models of an application domain that help to exchange and share knowledge with the help of ICT systems. “An ontology is a formal, explicit specification of a shared conceptualization” [Gruber 1993, 199]. Ontologies have to be *formal* which requires that they should be machine-readable. However, there are different degrees of formality of ontologies, from a thesaurus like WordNet to ontologies capturing formal theories for common-sense knowledge like Cyc. *Explicit specification* means that concepts and relationships as well as constraints on the use of concepts are defined openly and not left to the interpretation of the ontology’s users. *Shared* refers to the requirement that the conceptualizations made in an ontology have to be agreed upon. Finally, *conceptualization* is an abstract model, a representation of a domain or phenomenon which investigates concepts of that domain or phenomenon relevant to the ontology’s users.

Ontologies are not static, but evolve over time. Ontologies model objects in domains, relationships among those objects, properties, functions and processes involving the objects and constraints on and rules about the objects [Daconta et al. 2003, 190]. Thus, ontologies support clear-cut, concise, semantically rich and unambiguous communication between persons aided by an ICT environment. In the Semantic Web stack [Berners-Lee et al. 2001], the Web Ontology Language (OWL) helps to semantically map descriptions of resources serialized with the help of the Resource Description Format (RDF). OWL provides a standardized markup language that can be used to describe the top-level ontology that maps the meta-data standards described in Section 3.

3. Meta-data Standards

There are several high-level meta-data standards. One high-level schema is the Meta-data Encoding and Transmission Standard (METS) developed by the Digital Library Federation. METS proposes to define meta-data elements for complex digital objects. It offers a flexible mechanism for encoding descriptive, administrative, and structural meta-data for a digital library object, and for expressing the complex links between these various forms of meta-data. It is therefore a useful standard for exchanging digital library objects between repositories. In addition, METS provides the ability to associate a digital object with behaviours or services.

There are many initiatives for domain-specific content-oriented standardization. Knowledge work comprises active participation in processes of knowledge conversion, i.e. according to Nonaka and Takeuchi [1995] processes of externalization, socialization, internalization and combination. From the perspective of enterprise knowledge infrastructures, these processes are supported by publishing, discovery, collaboration, learning as well as integration and search services [Maier 2007, 318]. The underlying resources that support these services can be structured into personal information, personal communication, resource information, and specific information like text, images, audio and video which are examined in the following.

3.1. *Personal Information*

Personal information contains address books, calendars, task lists, wish lists, schedules, notes, etc. There are many software systems available that manage all or part of this information. In former days, paper-and-binder personal organizers have been used to administer personal information. These were small books containing a calendar, an address book and ruled pages for arbitrary notes. Nowadays, this information is held on our computers and synchronized with personal digital assistants (PDAs) to be conveniently available at any place.

Meta-data standards and specifications for personal information include vCard, vCalendar, FOAF, etc. vCard automates the exchange of personal information. It is a set of meta-data elements that covers and extends information typically found on a traditional business card [vCard]. vCalendar defines a transport and platform-independent format for exchanging calendaring and scheduling information in an easy, automated, and consistent manner [vCalendar]. It captures information about event and to-do items that are used by applications such as personal information managers and group schedulers. FOAF stands for *Friend Of A Friend* [FOAF]. It is based on XML and RDF and is intended to manage online communities. The FOAF namespace provides a set of properties and classes, focusing initially on people, documents, organizations, images.

3.2. *Personal Communication*

Communication is the process of exchanging information, e.g., in order to share knowledge and experiences. To a great extent, personal communication is done via e-mail messages. Other forms of personal communication include chat rooms, instant messaging, bulletin boards, and newsgroups. Personal communication may support voice and video.

An important standard for personal communication is the simple mail transfer protocol (SMTP or ESMTP), a text-based protocol for transferring messages. It is an application layer protocol that uses TCP/IP to transport and route its messages. SMTP uses commands like HELO or AUTH and status codes to communicate conditions between sender and receiver. SMTP is based on pure ASCII text. The standards describing mechanisms for transmitting text and other information are laid out in RFC2822, RFC2045, etc. Well-known elements of the RFC2822 standard include the destination address fields to, cc and bcc which specify the recipients of messages [RFC2822]. Originator fields contain the fields from, sender and reply-to. Identification fields contain a message identifier. Informational fields contain a subject and comments.

3.3. *Resource Descriptions*

The term electronic resource denotes a wide range of varying formats, such as (hyper) text documents, images, audio or video files. In spite of the format-specific differences between types of resources, a substantial part of their description can be dealt with in a generic, type-independent form. Standardization efforts consequently aim at designing a generic set of meta-data fields (Section 3.3.1) and standardized mark-up languages (Section 3.3.2) that are used to describe resources independent of their types and formats.

3.3.1. *General Fields*

The Dublin Core Meta-data Initiative (DCMI) began in 1995 with a workshop in Dublin, Ohio. Librarians, digital library researchers, content providers, and text mark-up experts started to improve discovery standards for information resources. The original Dublin Core (DC) emerged as a small set of descriptors drawing global interest from a variety of information providers in arts, sciences, education, business, and government [ISO 2003].

DC is intended to co-exist with meta-data standards that offer other semantics. It is expected that descriptive records will contain a mix of elements from various meta-data standards. DC's simplicity is both a strength and a weakness. Simplicity lowers the cost of creating meta-data and promotes interoperability. But it does not accommodate the semantic and functional richness supported by complex meta-data schemes. The use of richer meta-data schemes in combination with DC is encouraged. Richer schemes can also be mapped to DC for export or for cross-system searching. Conversely, DC records can be used as a starting point for the creation of more complex descriptions [ISO 2003].

3.3.2. *Mark-up Languages*

Text mark-up languages are used to tag content and turn unstructured and thus non-processable text into semi-structured text. Based on SGML, especially XHTML and XML have gained wide recognition as standard languages for exchanging data on the Web [XHTML, XML] or between application systems within and across organizations. With respect to resource description, mark-up languages provide a form of meta-language which can be used for partners exchanging resource descriptions to agree on the format in which meta-data is serialized and interpreted. XML has long been widely recognized as the most important base standard in this respect. Whereas Dublin Core standardizes the contents of a generic set of meta-data fields, mark-up languages standardize the format in which the standardized meta-data fields are serialized, in this case into XML elements. Using XML, a number of institutions have developed standards and started initiatives in order to provide comprehensive frameworks for definition and exchange of meta-data. The most prominent one is W3C's Semantic Web initiative.

3.4. *Specific Standards*

In addition to the generic standards, specific standards focus one or a limited number of specific types of resources. Successful standardization initiatives have been working on specific standards for texts, images and audio/video files. These are presented in the following subsections along the lines of a general description of the specific resource type and a brief description of widely known exemplary standards.

3.4.1. *Texts*

The advantage of text with respect to its treatment in solutions for semantic content management is that its content is fully accessible to text processing, searching and retrieving. In addition to meta-data on text, thus the content itself can be structured according to guidelines and standards so that not only the text document in its entirety can be accessed, but also individual parts of the document, e.g., a certain section, paragraph or embedded resources, such as figures, tables, formulae etc.

Examples for encoding schemes that focus on text documents are the Text Encoding Initiative's [TEI] encoding scheme or the Open Document Format maintained by the OASIS organization. The Open Document Format defines an XML schema for office documents, including, but not limited to text documents, spreadsheets, charts and graphical documents like drawings or presentations. Thus, text documents are not the only object targeted by the Open Document standard. The schema aims at high-level information suitable for editing and automatic processing of documents. The standard comprises pre-defined meta-data elements as part of a separate XML element usually located at the beginning of a document. Meta-data elements can be extended by users in a specified format requiring name, data type and value (user-defined meta-data) as well as in an unformatted way (custom meta-data). Pre-defined meta-data elements contain the elements defined in the Dublin Core standard [Durusau et al. 2006].

3.4.2. *Images*

Finding specific images is much harder than finding text, unless additional, textual, information is available. This additional information can be embedded in image files or stored in separate files. For example, users can upload their images to a server and then describe the contents of these images. Other users can then search these descriptions and find the corresponding images. Nowadays, most digital cameras embed various pieces of information into images they generate. This information includes current date/time, shooting conditions, e.g., whether a flash was used, GPS coordinates and camera settings, e.g., shutter, aperture. Adobe Photoshop, the leading image editing software, can also embed pieces of meta-data into image files, such as author, title or copyright notice.

Examples of image files containing meta-data include Exif (exchangeable image file format) and Tiff (tagged image file format). Exif is the format for camera-embedded meta-data [EXIF]. Exif is a standard for storing interchange information in image files, especially those using jpeg compression. Most digital cameras now use the Exif format. Tiff is a wide-spread format for high color depth images, besides jpeg and png. The Tiff format is supported by a wide range of applications [TIFF], e.g., Adobe Photoshop.

3.4.3. *Audio and Video*

Searching audio and video information is comparable to searching images. Without additional textual information it is hardly possible to create, e.g., multimedia directories or audio/video catalogues. Meta data specific to audio and video includes the original medium, e.g., reel-to-reel, sampling rate and bit depth, digitization date or digitization software used. Audio meta-data includes title, artist, track number, comment, year, genre or album title. Video meta-data includes title, directors, actors, format, language or studio.

Examples of formats include MPEG-7 and MP3. MPEG-7 is officially called "Multimedia Content Description Interface", a means of attaching meta-data to multimedia [Thom et al. 1998]. MPEG-1, -2, and -4 have been designed to represent information itself. MPEG-7 is meant to represent information about information, i.e., MPEG-7 supports finding multimedia content. MPEG-7 is supposed to make the web more searchable for multimedia content [MPEG-7]. Meta-data may be used for more advanced access to underlying content, by enabling automatic multimedia presentation or editing. MP3 stands for MPEG-1 Audio Layer-3 and is a standard technology and format for the compression of sound sequences into small files without losing too much sound quality [MP3]. Rippers and encoders can be used to convert songs from CDs to MP3 files. ID3 is the meta-data container that is most often used in conjunction with MP3. It allows information such as title, artist or album to be embedded in the MP3 file.

4. Top-Level Ontology

We suggest to use a top-level ontology and to merge the vast and increasing number of meta-data standards and ontologies under a single “roof”, the top-level ontology. We imagine six dimensions to be used for that purpose, i.e., the dimensions when, what, where, who, why and how. These dimensions have been identified as being essential for effective categorization, visualization and navigation of collections of contents in a series of unstructured interviews about and self-observation of ICT work environments of knowledge workers [Maier/Sametinger 2002]. We have used the terms isolation, loss and plurality in order to categorize problems faced with traditional ICT work environments based on the desktop metaphor. Problems arise due to the facts that we have (1) a hierarchical folder structure on our desktop rather than a flexible means of categorization, (2) trivial and multiple categorization mechanisms in various applications, e.g., Lotus Notes, Mozilla Firefox, MS Exchange, MS Internet Explorer, Palm Desktop, (3) documented knowledge that is only accessible via specific applications, e.g., email messages, appointments including notes, ideas and protocols or information about people's skills and expertise and their current availability, (4) versioning information only available with specific applications, e.g., DMS, (5) multiple documents that are different representations of the same contents, and (6) insufficient meta-data about local and remote documents. Thus, we suggest mapping entries of meta-data standards provided by a variety of applications and adhering to a variety of standards to the identified six basic dimensions therefore easing retention, discovery and presentation of documented knowledge.

Meta-data standard items map differently to our top-level ontology, see Fig. 1. Exif items, for example, are mostly mapped to the how-dimension. This is not surprising because when we take a picture with our digital camera, then most of the information that is available about the taken image is technical information about its creation like shutter speed o aperture opening. ID3 tags also provide a bulk of technical information. Additionally, they contain information about persons like composer and interpret.

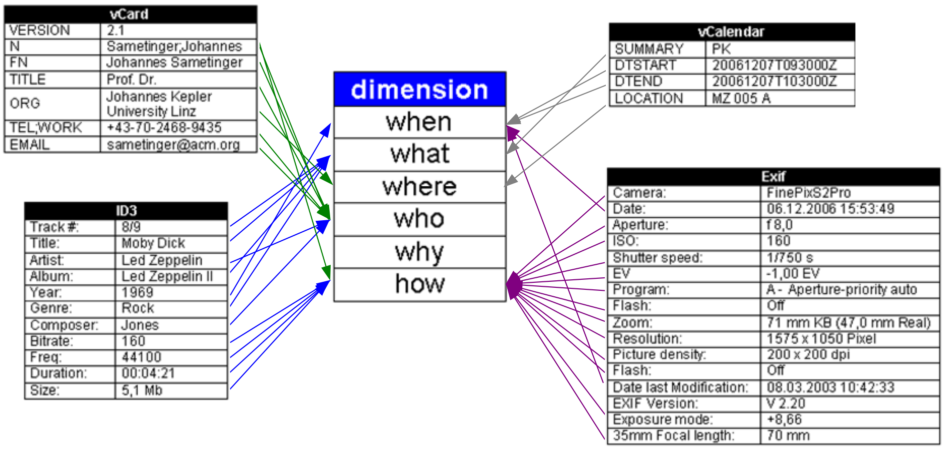

Fig. 1. Mapping of meta-data standard items to top-level ontology

There is one thing that catches our eye when looking at the suggested matching. It is the fact that there are mappings for all dimensions except for one, i.e., the why-dimension. What seems odd at first sight becomes obvious when giving it a second thought. Information in this dimension cannot be created automatically but requires input from a human being. For example, a picture may have been taken because there was the wedding of a relative. The camera will never know this. We will have to add this extra piece of information in order to keep it and make it accessible in the future. The subject of an email message is an example for information that we map to the why-dimension.

The top-level ontology relates meta-data elements to the six dimensions. In the following, a case example describes how these mappings can be used in order to respond to rich user queries. A knowledge worker is engaged in an on-demand learning activity and needs to find all relevant resources that could answer questions regarding the topic “knowledge management”. The typical entry point in this case would be the what-dimension. The knowledge worker would find events in the organization’s group calendar (meta-data element summary), electronic text documents that contain documented knowledge on the topic (title, keywords), audio and video presentations with systematic introductions into the topic (title, genre=learning) as well as personal information on co-workers whose job title has to do with knowledge management. Browsing the material, the knowledge worker realizes that “John Brown” seems to be the organization’s top expert on the topic, so she initiates another query using the who-dimension in order to find out contact details, events on which she could personally meet up with the expert as well as text, audio and video documents authored by or featuring the expert. The results are presented in one or two dimensions. One single dimension can be selected twice, e.g., a who/who-matrix that shows the relationships between our knowledge worker and the expert. In the intersection cell, all resources are reported that contain meta-data on “Sue Ash” AND “John Brown”, e.g., email messages, a document created by John and modified by Sue (single resources are thus reported as often as retrieved via multiple paths), a meeting in which both participated or friends and business acquaintances that they have in common (using the friend-of-friend framework).

In an empirical study of the TOP 500 German companies conducted in 1999/2000, *improving visibility of knowledge* and *improving access to existing knowledge* were the two most important goals that organizations strive for when implementing KM initiatives [Maier 2007, 472]. Our simple approach can be seen as targeting these two goals which have also been proposed as starting points for KM life cyles [e.g., Probst et al. 1998]. In addition, our approach positively influences the effort needed to *improve retention of knowledge* in a variety of forms and formats and to *improve acquisition of external knowledge* which is already externalized and available in a documented form. These are another two goals found less important by organizations in the cited study, but which have been on the rise since then with a recent stronger focus on the management of knowledge risks, e.g., concerning intra-, inter- and extra-fluctuation of employees, as well as on projects, processes and formal as well as informal knowledge cooperations crossing organizational boundaries [Maier 2007, 136]. The challenge we target with respect to these goals is that documented knowledge embedded in varying intra- and inter-organizational contexts needs to be jointly retained and accessed while the effort to prepare and refine the documented knowledge should be kept as low as possible.

5. Comparable work

Related work includes desktop alternatives like Lifestreams [Fertig 1996], Time-Machine Computing [Rekimoto 1999] or Presto [Dourish 1999], and tools to display large amounts of information, e.g., TheBrain (www.thebrain.com), Data Mountain [Robertson 1998], DLITE [Cousins 1997], and TopicShop [Amento 2000]. Currently, information on our computers is still organized in old-fashioned files and folders, but searched by specialized tools like Google Desktop Search [Google] or Copernic Desktop Search [Copernic]. Operating systems also have features to better handle large amounts of data. This is done by adding search capabilities like Apple's Spotlight [Apple] or Windows Vista's Instant Search [Microsoft]. Tools like Google's Picasa [Picasa] and Apple's iTunes [iTunes] provide interesting approaches for organizing large amounts of data with the drawback, that extra software has to be used which is not integrated into the user's desktop.

6. Conclusion

We argue that our computers would greatly benefit if their access mechanisms would be extended to use a top-level ontology and to have dimensional views on documents and electronic resources. In addition, the benefit would be much greater if all resources on our computers would share such an access mechanism. This means that we do not need any longer, for example, applications administrating images and songs or, in a knowledge management perspective, lessons learned, learning objects, contributions to newsgroups, Wikis and Weblogs, personal experiences and the like. Rather, we would have a simple file access mechanism with additional small tools that provide the functionality that applications now offer in addition to administering file access.

This paper has shown that existing meta-data standards can be mapped to a simple six-dimensional top-level ontology which enables convenient integration between a variety of heterogeneous types of resources and their specific meta-data standards. The main advantage of this approach is its simplicity that does not require the time-consuming mapping of sophisticated ontologies and the implementation of costly ontology management middleware. Access to electronic resources fostered by the use of a top-level ontology thus provides an easy mechanism to enhance the ICT work environment of knowledge workers and thus the support for weakly structured, highly complex, knowledge-intensive business processes with their unpredictable knowledge requirements.

References

Amento, A., Terveen, L., Hill, W., Deborah Hix, D. (2000): TopicShop, in: Proceedings of the 13th annual ACM symposium on User interface software and technology. November 2000

Apple: Tiger Developer Overview Series: Working with Spotlight. http://developer.apple.com/macosx/spotlight.html, last visited May 22, 2007

BernersLee, T., Hendler, J., Lassila, O. (2001): The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, May 17th, 2001, http://www.sciam.com/

Carr, L., Miles-Board, T., Woukeu, A., Wills, G., Hall, W. (2004): The Case for Explicit Knowledge in Documents, in: Proceedings of the 2004 ACM Symposium on Document Engineering, New York, 2004, 90-98

Copernic: Copernic Desktop Search, http://www.copernic.com/en/products/desktop-search/index.html, last visited May 22, 2007

Cousins, S., Paepcke, A., Winograd, T., Bier, E. A., Pier, K. (1997): The Digital Library Integrated Task Environment (DLITE), in: Proceedings of the 2nd ACM international conference on Digital libraries DL 97, Philadelphia PA, USA, July 1997

Daconta, M. C., Obrst, L. J., Smith, K. T. (2003): The Semantic Web. A Guide to the Future of XML, Web Services and Knowledge Management, Indianapolis (IN), USA 2003

Dourish, P., W Edwards, W., K., LaMarca, A., Salisbury, M. (1999): Presto: an experimental architecture for fluid interactive document spaces, in: ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 6, No. 2, June 1999

Durusau, P., Brauer, M., Oppermann, L. (2006): Open Document Format for Office Applications (OpenDocument) v1.1, Committee Specification1, 19 Oct 2006, http://www.oasis-open.org/committees/office, last visited March 12, 2007

EXIF: JEITA. Exchangeable image file format for digital still cameras: Exif Version 2.2, http://www.exif.org/Exif2-2.PDF, April 2002. last visited May 22, 2007

Fertig, S., Freeman, E., Gelernter, D. (1996): Lifestreams: An Alternative to the Desktop Metaphor, in: Proceedings of the CHI '96 conference companion on Human factors in computing systems: common ground, Vancouver (BC), Canada, April 1996

FOAF. The Foaf Project. http://www.foaf-project.org/

Google: Google Desktop Search, http://desktop.google.com/, last visited May 22, 2007

Gruber, T. R. (1993): A Translation Approach to Portable Ontology Specifications, in: Knowledge Acquisition, Vol. 5, No. 2, 1993, 199-220

ISO (2003): Information and documentation — The Dublin Core meta-data element set. ISO Standard 15836-2003: http://www.niso.org/international/SC4/n515.pdf, February 2003, last visited March 12, 2007

iTunes: Apple’s iTunes http://www.apple.com/itunes/overview/, last visited May 22, 2007

Maier, R. (2007): Knowledge Management Systems. Information and Communication Technologies for Knowledge Management, 3rd edition, Berlin

Maier, R., Hädrich, T., Peinl, R. (2005): Enterprise Knowledge Infrastructures, Berlin

Maier, R., Peinl, R. (2005): Semantic Description of Documents in Enterprise Knowledge Infrastructures, in: Jantke, K. P., Fähnrich, K-P., Wittig, W. S. (eds.): Marktplatz Internet: Von e-Learning bis e-Payment. Proceedings of the 13th Leipziger Informatik-Tage, LIT 2005, September, 21-23, 2005, Leipzig, 358-366

Maier, R., Sametinger, J. (2002): Infotop - An Information and Communication Infrastructure for Knowledge Work, ECKM2002, The 3rd European Conference on Knowledge Management, Trinity College Dublin, Ireland, September 24-25, 2002

MP3: MP3: ISO/IEC JTC1/SC29/WG11 N, MPEG-1 - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, June 1996. http://www.chiariglione.org/mpeg/standards/mpeg-1/mpeg-1.htm. last visited May 22, 2007

MPEG-7: ISO/IEC JTC1/SC29/WG11N6828, MPEG-7 Overview, http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.. last visited May 22, 2007

Nonaka, I., Takeuchi, H. (1995): The Knowledge Creating Company. New York

Picasa-Google's photo organizer, Picasa, http://picasa.google.com/, last visited May 22, 2007

Probst, G., Raub, S., Romhardt, K. (1998): Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen, 2nd edition, Wiesbaden

Rekimoto, J. (1999): Time-machine computing: a time-centric approach for the information environment, in: Proceedings of the 12th annual ACM symposium on User interface software and technology, Asheville, North Carolina

RFC2822 (2001): Internet Message Format, Network Working Group, Request for Comments: 2822, http://www.ietf.org/rfc/rfc2822.txt, last visited May 22, 2007

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., van Dantzich, M. (1998): Data mountain, in: Proceedings of the 11th annual ACM symposium on User interface software and technology, November 1998

Schimkat, R.-D. (2003): Techniken und Aspekte zur Realisierung proaktiver Informationssysteme. Osnabrück

Swetland A. G., Baca M. (Eds.) (2000): Introduction to Meta-data: Pathways to Digital Information, Getty Information Institute

TEI. Text Encoding Initiative (2007), http://www.tei-c.org/ last visited March 12, 2007

Thom D., Purnhagen H., and the MPEG Audio Subgroup (1998): MPEG Audio FAQ Version 9 MPEG-7, October 1998; see http://www.chiariglione.org/mpeg/faq/mp7-aud/mp7-aud.htm, last visited May 9, 2005

TIFF: Adobe. TIFF 6.0 Specification, http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf, last visited May 22, 2007

vCalendar: http://www.imc.org/pdi/vcaloverview.html, last visited December 5, 2005

vCard: Version 2.1, http://www.imc.org/pdi/vcard-21.rtf, last visited December 5, 2005

Wolff, E. N. (2005): The Growth of Information Workers, in: Communications of the ACM, Vol. 48, No. 10, 37-42

Microsoft: Features Explained: Instant Search, http://www.microsoft.com/windows/products/windowsvista/features/details/instantsearch.mspx, last visited May 22, 2007

XHTML: XHTML 1.0, The Extensible HyperText Markup Language, 2nd edition, A Reformulation of HTML 4 in XML 1.0, W3C Recommendation, 26 January 2000, revised 1 August 2002, last visited May 22, 2007

XML: Extensible Markup Language (XML) 1.0, 4th edition, W3C Recommendation, 16 August 2006, edited in place 29 September 2006, last visited May 22, 2007